AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 101:151–159 (1996)

Assessing the Reliability of Paleodemographic Fertility Estimators Using Simulated Skeletal Distributions

RICHARD R. PAINE AND HENRY C. HARPENDING
Department of Anthropology, University of Utah, Salt Lake City, Utah 84113 (R.R.P.) and Department of Anthropology, The Pennsylvania State University, University Park, Pennsylvania 16802 (H.C.H.)

KEY WORDS Demographic reconstruction, Age-at-death distributions, Model life tables

ABSTRACT The reliability of published paleodemographic fertility reconstruction methods was assessed using simulated age-at-death distributions and a published cemetery series from a population with known birth rates. In the first test, the Brass ([1971] Biological Aspects of Demography, pp. 69–110) LOGIT models were used to generate 180 simulated skeletal samples of various sizes (N = 50, 100, 250) from hypothetical populations with known demographic rates. The base populations were expanding (r = 0.01), stationary, or declining (r = −0.01), yet all had the same life expectancy. Growth differences resulted from different fertility rates. The simulated skeletal series were then analyzed using the model life table fitting procedure outlined by Paine ([1989a] Am. J. Phys. Anthropol. *79*:51–62), three commonly employed age ratio tests (Bocquet-Appel and Masset [1892] J. Hum. Evol. *11*:321–333; Buikstra et al. [1986] Am. Antiquity *51*:528–546), and one age-at-death ratio not previously published. In the second test the model life table fitting procedure was used to estimate fertility for a historical population, the Newton Plantation, Barbados (Corruccini et al. [1989] Am. Antiquity *54*:609–614), with known demographic characteristics. © 1996 Wiley-Liss, Inc.

Documenting changes in human demographic characteristics is an essential element in the study of important developments, including the adoption of agriculture and the effects of greater sedentism and urbanism on human health (Boserup, 1965; Cohen, 1977, 1989; Cohen and Armelagos, 1984). Fertility is the demographic characteristic that can be reconstructed most reliably (Johansson and Horowitz, 1986; Konigsberg and Frankenberg, 1994; Milner et al., 1989; Paine, 1989a; Sattenspiel and Harpending, 1983) the reliability of methods for reconstructing fertility rates from skeletal age-at-death distributions has received considerable attention and criticism (e.g., Bocquet-Appel and Masset, 1982, 1985; Buikstra and Konigsberg, 1985; Buikstra et al., 1986; Corruccini et al., 1989; Horowitz and Armelagos, 1988).

In this paper we assess and compare the reliability of several published methods for estimating fertility rates from age-at-death distributions. We do not consider the reliability of the age estimation techniques themselves. For purposes of analysis, we make the (perhaps unrealistic) assumption that the age estimates we are dealing with are accurate—at least within broad categories. There has been considerable discussion of age estimation techniques (e.g., Bocquet-Appel and Masset, 1982; Konigsberg and Frankenberg, 1992; Konigsberg et al., in press; Lovejoy et al., 1985; Masset, 1989; Meindl and Lovejoy, 1989). Though we consider age estimation a

Received November 21, 1995; accepted May 3, 1996.

serious problem, we consider it a distinct problem to be addressed elsewhere.

DATA AND METHODS

Test 1: Simulated distributions

We generated 180 simulated skeletal samples of various sizes (N = 50, 100, 250) from three hypothetical populations with known demographic rates. These sample sizes are fairly typical of most archaeological skeletal collections. The base populations were expanding (r = 0.01; CBR = 31), stationary (r = 0; CBR = 23), or declining (r = −0.01; CBR = 16), yet all had the same life expectancy. Growth differences resulted from different fertility rates. To counter objections concerning the applicability of the Coale and Demeny model west patterns to paleodemographic populations, and to make the simulated series independent of analysis, the Brass (1971) LOGIT models were used to generate the simulated series.

The Brass LOGIT system for generating model life tables is based on a generic human survival function given by Brass as values of 1_x at ages 1, 5, 10, 15, . . ., 100 (Newell, 1988). The expectation of life at birth for Brass' standard survival function is about 50 years (Newell, 1988).

The function can be transformed by a LOGIT transformation into any of a family of curves by varying either of two parameters. The first parameter measures overall mortality. The second measures the ratio of mortality early in the lifespan to mortality late in the lifespan. Brass shows that this set of curves provides a good description of the mortality experience of human populations over a wide range of levels of overall mortality. We used the Brass method for computational convenience, and to maintain independence from the model west-based procedure being tested, but it is our experience that using Brass models, the Coale and Demeny models, or any other set of model tables makes little or no difference to our results.

Since we wanted to simulate the proportions of deaths that a population would leave at the very ages it would leave them, we needed some further manipulation to generate our model death distributions. We interpolated through the survival function values given by Brass with a spline so we would have year by year values. The successive differences of these values give the risk of death by year for a newborn. The risk of a person age x dying before age x + 1 is given by:

$$q_x - (1_x - 1_{x+1})/1_x.$$

It is well known that the age distribution of the living in a stable population is given by:

$$p_x = e^{-rx}\, 1_x.$$

So the expected number of deaths is proportional to r p.

Given the distribution of expected ages at death we generated simulated skeletal samples by drawing uniformly distributed random numbers and binning them according to the cumulative distribution of skeletal age. For example, if the probability of dying before age 40 is 0.5 and the probability of dying before age 50 is 0.6, then any random number drawn that is between 0.5 and 0.6 contributes a death to the age category 40–49. The distribution of simulated skeletal series is presented in Table 1.

TABLE 1. Simulated skeletal series by growth rate and sample size

	Intrinsic growth			
Sample size	r = 0.01	r = 0.00	r = −0.01	total
50	20	20	20	60
100	20	20	20	60
250	20	20	20	60
Total	60	60	60	180

Analysis

The simulated skeletal series from these populations were subjected to analysis using four methods: the model life table fitting procedure developed by the authors and outlined by Paine (1989a); two commonly employed age ratio tests (Bocquet-Appel and Masset, 1982; Buikstra et al., 1986); and another age-at-death ratio suggested by Milner et al. (1989:52). Comparative evaluation of the various methods was based on two criteria: (1) the ratio of the standard deviation of the estimates to the mean birth rate estimate (or the value of the age ratio under each of the formulae—Bouquet-Appel and

Masset, 1982; Buikstra et al., 1986) for each sample size and growth rate subsample; (2) the ability of each method to differentiate expanding, stationary, and declining populations. This ability is tested by attempting to separate simulated skeletal series into the three subgroups with different growth rates.

Translating any of the three age-at-death ratios into an actual fertility estimate requires regressing the ratio against a given set of life table models. Conventionally they have been regressed against the Coale and Demeny (1966, 1983) west model tables. However, these age-at-death ratios can be regressed against any given set of life table models at the discretion of the investigator. For example, Sullivan (in press) has recommended using the Coale and Demeny "north" models to analyze Contact Period Huron ossuaries. He bases this recommendation on local evidence of tuberculosis. Where site- or region-specific information can be used to reinforce model choice, investigators should be able to adjust analyses accordingly. Therefore, no specific fertility estimates, which would have required the imposition of a specific model, were attempted. We have broadened the second criterion to test each method's ability to differentiate a declining, stationary, or expanding population, rather than its ability to estimate specific fertility or birth rates.

Coale and Demeny's west model life tables are specified in the model life table fitting procedure (Paine 1989a—the authors have written alternative programs that employ the other Coale and Demeny model life table families). The testing was taken one step further for this method. Crude birth rate estimates (CBR) from the model life table fitting procedure were compared to the "correct" values of the simulated populations. Gross reproduction rate estimates, the manipulated fertility variable in the procedure, were carried to one decimal place only, as were CBR estimates (see Paine, 1989a,b).

RESULTS

The model life table fitting procedure is an effective means of reconstructing fertility levels, according to all three criteria outlined above. Standard deviations for all simulation classes, including those with a sample size of 50, are less than 25% of the sample mean. The correct birth rate falls within one standard deviation of the mean of the estimates for every sample group, including those with a sample size of 50. The mean distance between the mean birth estimates and the actual population birth rates is only 2.95%. The 95% confidence intervals for the three growth classes are fully distinct for all sample sizes (Table 2), though Figures 1, 2, and 3 show that there is limited overlap of estimates for series from adjacent population types with samples of 50. This overlap is virtually nonexistent in the analyses of series of 250 individuals. Fertility estimates based on this procedure are affected by infant underenumeration and the patterns of adult age estimation bias described by Lovejoy et al. (1985) (Paine, 1992; Paine and Harpending, in preperation).

TABLE 2. Mean and variance statistics for birth rate estimates at given actual fertility rates and simulated distribution sample sizes

Sample size	Mean	Standard deviation	s/x	95% CI (Student's t)
Model crude birth rate (CBR) = 16; intrinsic growth (r) = −0.01				
50	16.40	4.07	0.248	(14.50–18.31)
100	15.55	2.37	0.153	(14.44–16.66)
259	14.65	2.11	0.144	(13.66–15.64)
Model crude birth rate (CBR) = 23; intrinsic growth (r) = 0.00				
50	23.00	4.24	0.184	(21.01–24.99)
100	23.50	3.27	0.139	(21.97–25.03)
250	22.35	1.63	0.073	(21.59–23.11)
Model crude birth rate (CBR) = 31; intrinsic growth (r) = +0.01				
50	31.65	5.79	0.183	(28.94–34.36)
100	30.75	3.45	0.112	(29.14–32.36)
250	30.10	2.38	0.078	(29.07–31.19)

The Bocquet-Appel and Masset D5-14/D20+ (1982) formula is not an effective method when sample size is 100 or less (Table 3). It fails by either criterion outlined for the test. Especially in stationary or declining populations, there simply are not enough deaths between the ages of 5 and 14 years to overcome stochastic variation. Sample

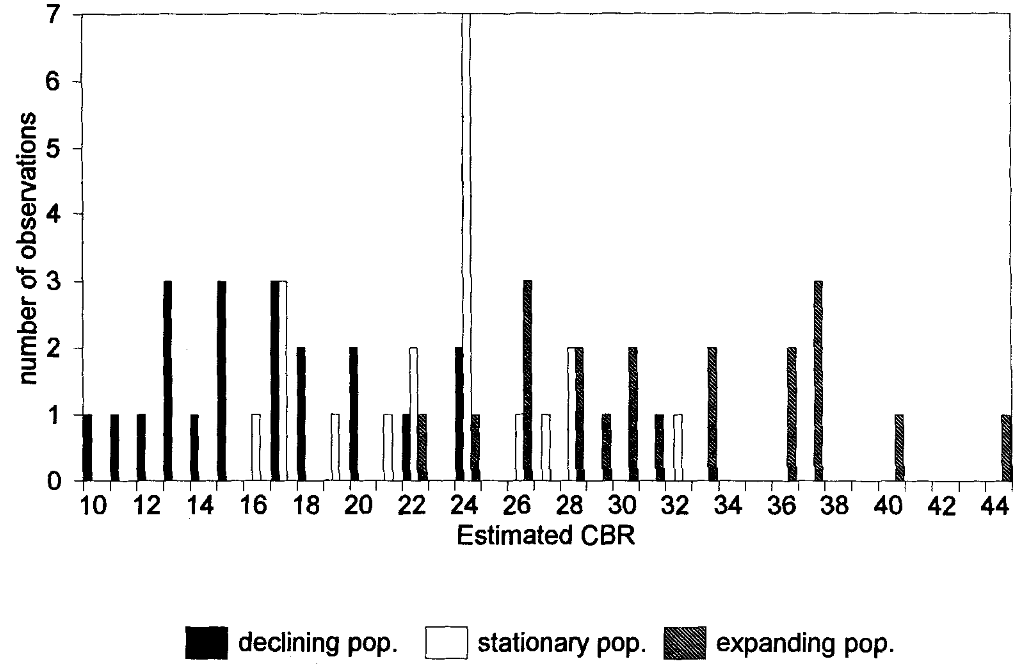

Fig. 1. Distribution of maximum likelihood–based birth rate estimates for simulated skeletal series with N = 50.

size at these ages is very small and there are numerous empty cell values at points crucial to the analysis.

The Buikstra et al. (1986) ratio of deaths beyond age 30 years to all deaths beyond age 5 years (D30+/D5+) has the lowest ratio of standard deviation to sample mean, averaging less than 5%. However, the method fails to differentiate clearly among the three fertility/growth classes, though the mean values for the three classes visibly displayed the expected pattern. The 95% confidence intervals for the declining and stationary models (CBR 16 and 23, respectively) are not statistically distinct.

Buikstra et al.'s (1986) second ratio, deaths to individuals between 1 and 5 years old to deaths to individuals between 1 and 10 years old (D1–5/D1–10), is less reliable. It is affected by small sample size, in the age classes employed, in much the same manner as the Bocquet-Appel and Masset (1982) formula. It was not possible to differentiate series from the three model populations with this ratio.

The ratio of deaths below age 45 to those above age 45 clearly differentiates the three classes of simulated distributions. The 95% confidence intervals of the sample distributions are distinct, and the standard deviation is, on average, less than 20% of the mean value of the ratio. This ratio is significantly affected by infant underenumeration and by age estimation bias (Harpending and Paine, 1992; Paine, 1992; Paine and Harpending, in preparation).

Test 2: The Newton Slave Plantation, Barbados: A test using age-at-death data from a population with known birth rates

Corruccini et al. assessed the relative effectiveness of fertility reconstruction methods in a comment to *American Antiquity,* using a slave population from Barbados with historically observed birth rates (Corruccini et al., 1989). They analyzed both the histori-

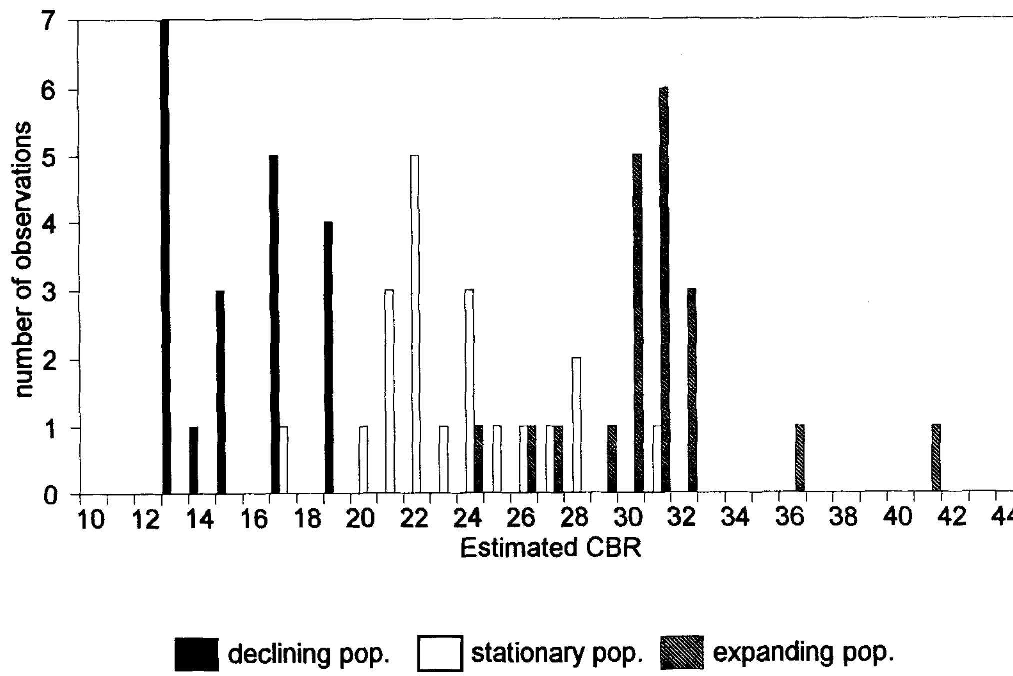

Fig. 2. Distribution of maximum likelihood–based birth rate estimates for simulated skeletal series with N = 100.

cally recorded ages at death and the skeletal record from the Newton Plantation, Barbados slave population using the Bocquet-Appel and Masset (1982) and the Buikstra et al. (1986) formulae. They concluded that the available fertility reconstruction methods (specifically the D5–14/D20+, and D30+/D5+ ratios) were not effective estimators of past birth rates (Corruccini et al., 1989: 611–612).

The life table fitting procedure presented here (and by Paine 1989a) was not available to Corruccini et al. (1989) when they wrote their comment. The historically observed deaths from their original studies (Corruccini et al., 1982; Handler and Lange, 1978) were analyzed to test the life table fitting procedure (Fig. 4). The Corruccini et al. (1989) paper shows that the skeletal sample differs markedly from the historically observed deaths, but the skeletal age distribution was not available for analysis. Table 4 is a reproduction of Corruccini et al.'s (1989:611) results for the historically observed deaths with the addition of a crude birth rate estimate based on fitting a model west life table. The life table fitting procedure provides the best estimate of the historically observed birth rates (Table 4). The magnitude of error is consistent with error observed in analyzing unbiased, simulated distributions of similar sample size.

CONCLUSIONS

When selecting an estimating procedure to reconstruct prehistoric fertility from skeletal series, the investigator is faced with two competing challenges. The first is to overcome sample bias. Many (most) archaeologically recovered skeletal series display age-at-death distributions that differ markedly from contemporary and historical distributions of death. They contain relatively fewer infants, young children, and advanced-age adults than expected (Howell, 1976; Konigsberg et al., in press; Paine, 1989b, in press; Weiss, 1973). Though this may reflect cultural (e.g., Zimmerman et al., 1981) or biological phenomena (see Keckler, in press;

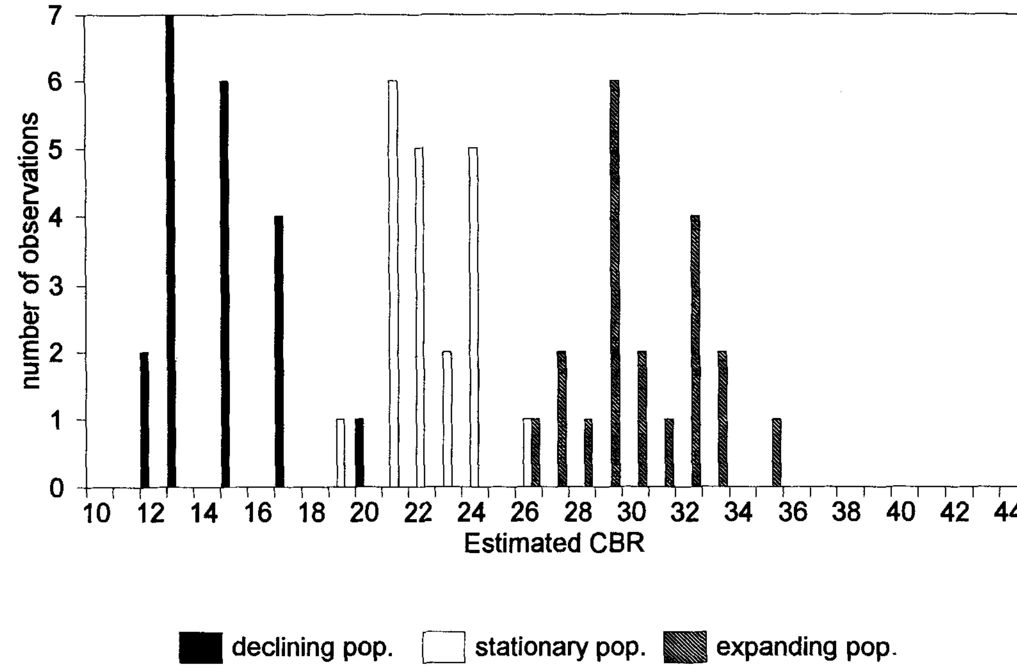

Fig. 3. Distribution of maximum likelihood-based birth rate estimates for simulated skeletal series with N = 250.

TABLE 3. Mean and variance statistics of several fertility estimation methods based on simulated populations (N = 100)

	Mean	Standard deviation	Coefficient of variation	95% CI (Student's *t*)
D5–14/D20+[1]				
CBR = 16	0.022	0.025	1.157	(0.0101–0.0339)
CBR = 23	0.056	0.057	0.440	(0.0441–0.0670)
CBR = 31	0.079	0.041	0.525	(0.0596–0.0984)
D30+/D5+[2]				
CBR = 16	0.892	0.037	0.042	(0.8744–0.9095)
CBR = 23	0.864	0.030	0.035	(0.8500–0.8784)
CBR = 31	0.791	0.049	0.062	(0.7674–0.8135)
D10–5/D1–10[2]				
CBR = 16	0.844	0.197	0.234	(0.7514–0.9362)
CBR = 23	0.778	0.220	0.283	(0.6753–0.8813)
CBR = 31	0.838	0.119	0.142	(0.7818–0.8933)
D0–44/D45+				
CBR = 16	0.488	0.102	0.209	(0.4399–0.5356)
CBR = 23	0.750	0.029	0.038	(0.6893–0.8101)
CBR = 31	1.278	0.235	0.184	(1.1687–1.3876)

[1] Bocquet-Appel and Masset (1982).
[2] Buikstra et al. (1986).

Lovejoy et al., 1977) in some cases, in many—if not most—it probably reflects bias in archaeological recovery (Walker et al., 1988) and osteological analysis (Lovejoy et al., 1985). Such bias can have dramatic effects on paleodemographic reconstructions (Harpending and Paine, 1992; Paine, 1992; Paine and Harpending, in preparation). Such bias is generally dealt with by employing age-at-death ratios (e.g., Bocquet-

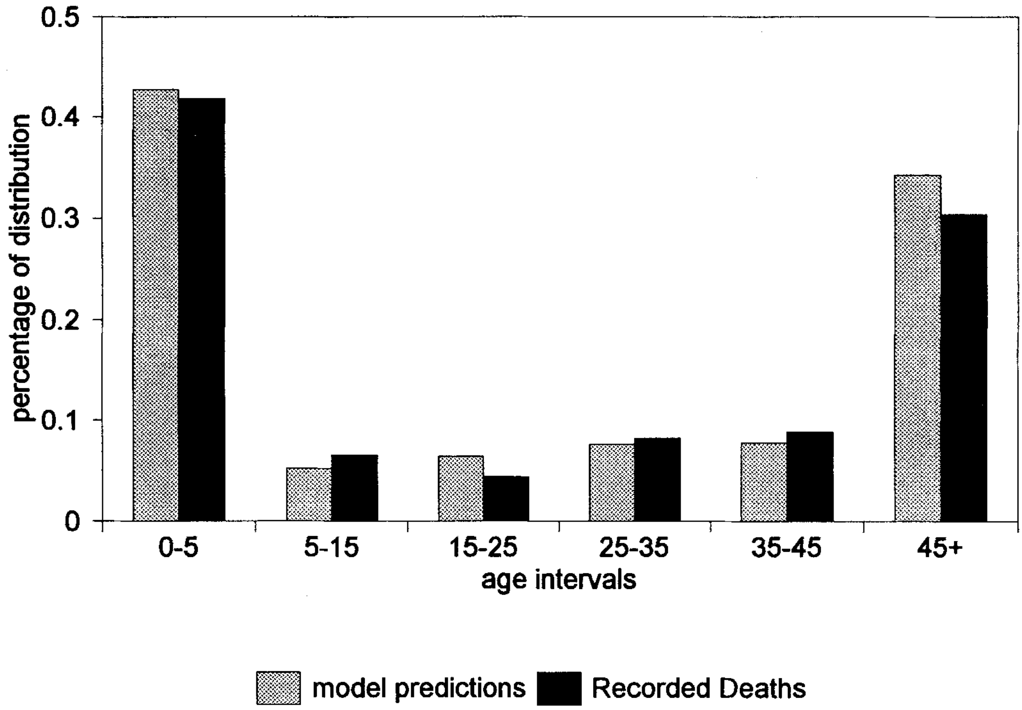

Fig. 4. Newton Plantation, Barbados (1776–1801): best-fitting model west distribution (maximum likelihood estimate) vs. recorded deaths. Recorded deaths are from Corruccini et al. (1982) and Handler et al. (1978).

TABLE 4. Results of Corruccini et al. (1989: 611) study with the addition of estimates based on model west life tables fitted by maximum likelihood estimation

Estimation method	CBR
Historically observed	37.0
D5–14/D20+[1]	32.1
D30+/D5+[2]	52.3
Fitted model west[3]	38.4

[1] Bocquet-Appel and Masset (1982); result as reported in Corruccini et al. (1989:611).
[2] Buikstra et al. (1986); result as reported in Corruccini et al. (1989:611).
[3] Paine (1989). All three estimates are based on historically observed deaths.

Appel and Masset, 1982; Buikstra et al., 1986) that do not rely on age classes prone to underenumeration. This can work well, given sufficiently large study populations, but can exacerbate the second area of challenge.

Small populations (unbiased, living populations) are subject to stochastic variation in both living structure and patterns of death. The impact of stochastic processes varies inversely with population size. Sampling introduces another level of similar bias. For example, in the 20 simulated distributions with $N = 100$ and $r = 0$, the number of deaths to individuals between 5 and 14 years of age varies from 0 to 7. Eliminating age classes from analyses in order to counter age bias can result in a crucial loss of data. This is demonstrated by the present analyses. Fertility reconstruction methods that eliminate large classes of data (individuals from 0 to 5 years and 40+ constitute the largest expected age classes in most demographic situations) are ineffective for small samples. The vast majority of archaeologically recovered skeletal series fall under the heading of "small samples."

The Corruccini et al. (1989) finding for Newton Plantation is consistent with those

from the simulated sample analyses presented above. The number of recorded deaths for the Newton plantation is 232. Of these, only 15 were between the ages of 5 and 14 years. Stochastic variation in such a small sample probably accounts for the lack of effectiveness of the Bocquet-Appel and Masset (1982) ratio. Analyses of the simulated distributions indicate that the D30+/D5+ ratio is relatively insensitive to varying fertility levels. The regressions to estimate crude birth rates from the Bocquet-Appel and Masset (1982) and the Buikstra et al. (1986) ratios shown in Table 4 were performed by Corruccini and coworkers, and are reproduced from their 1989 comment (Corruccini et al., 1989:611).

Largely because it employs all available data, the life table fitting procedure (Milner et al., 1989; Paine, 1989a) provides more accurate fertility estimates than the other widely relied upon estimators considered here for *small, unbiased* series. However, the procedure is prone to bias from infant and advanced adult underenumeration common to archaeologically recovered skeletal series. The degree to which it is affected is the subject of a forthcoming paper (Paine and Harpending, in preparation). When choosing a fertility estimator, investigators must weigh these two problems in relation to their specific skeletal series.

ACKNOWLEDGMENTS

We thank George R. Milner, James W. Wood, Lyle W. Konigsberg, Jennifer Graves, Valerie S. Gates, and the two anonymous reviewers for comments on earlier versions of this paper. Much of the work for this paper was completed while R.R.P. was a Charles P. Taft Postdoctoral Fellow at the University of Cincinnati. We thank the trustees of the Taft Foundation and the Department of Anthropology at the University of Cincinnati for their support and hospitality.

LITERATURE CITED

Bocquet-Appel JP, and Masset C (1982) Farewell to paleodemography. J. Hum Evol. *11*:321–333.

Bocquet-Appel, JP and Masset C (1985) Paleodemography: Resurrection or ghost. J. Hum Evol. *14*:107–111.

Boserup E (1965) The Conditions of Agricultural Growth: The Economics of Agrarian Change Under Population Pressure. New York: Aldine.

Brass W (1971) On the scale of mortality. In W Brass (ed.): Biological Aspects of Demography. London: Taylor and Francis, pp. 69–110.

Buikstra JE, and Konigsberg LW (1985) Paleodemography: Critiques and controversies. Am. Anthropol. *87*:316–333.

Buikstra JE, Konigsberg LW, and Bullington J (1986) Fertility and the development of agriculture in the prehistoric midwest. Am. Antiquity *51*:528–546.

Coale AJ, and Demeny P (1966) Regional Model Lifetables and Stable Populations. Princeton, NJ: Princeton University Press.

Coale AJ, and Demeny P (1983) Regional Model Lifetables and Stable Populations. New York: Academic Press.

Cohen MN (1977) The Food Crisis in Prehistory. New Haven: Yale University Press.

Cohen MN (1989) Health and the Rise of Civilization. New Haven: Yale University Press.

Cohen MN, and Armelagos GJ (eds.) (1984) Paleopathology at the Origins of Agriculture. New York: Academic Press.

Corruccini RS, Handler JS, Mutaw RJ, and Lange FW (1982) Osteology of a slave burial population from Barbados, West Indies. Am. J. Phys. Anthropol. *59*: 443–459.

Corruccini RS, Handler JS, Mutaw RJ, and Lange FW (1989) Inferring fertility from relative mortality in historically controlled cemetery remains from Barbados. Am. Antiquity *54*:609–614.

Handler JS, and Lange FW (1978) Plantation Slavery in Barbados: An Archaeological and Historical Investigation. Cambridge, MA: Harvard University Press.

Harpending HC, and Paine RR (1992) Measuring the accuracy of fertility reconstructions using simulated death distributions. Poster presented at the 61st annual meeting of the American Association of Physical Anthropologists, Las Vegas.

Horowitz S, and Armelagos GJ (1988) On generating birth rates from skeletal populations. Am. J. Phys. Anthropol. *76*:189–196.

Howell N (1976) Toward a uniformitarian theory of paleodemography. In KM Weiss and PE Smouse (eds.): The Demographic Evolution of Human Populations. New York: Academic Press, pp. 25–44.

Johansson SR, and Horowitz S (1986) Estimating mortality in skeletal populations; influence of the growth rate on the interpretation of levels and trends during the transition to agriculture. Am. J. Phys. Anthropol. *71*:233–250.

Keckler CNW (in press) Where did all the humans go? Simulating forager age-at-death distributions with a biphasic model of mortality. In RR Paine (ed.): Integrating Archaeological Demography: Multidisciplinary Approaches to Prehistoric Population. Carbondale: Center for Archaeological Investigations Occasional Papers No. 24.

Konigsberg LW, and Frankenberg SR (1992) Estimation of age structure in anthropological demography. Am. J. Phys. Anthropol. *89*:236–256.

Konigsberg LW, and Frankenberg SR (1994) Paleodemography: "Not quite dead." Evol. Anthropol. *3*: 92–105.

Konigsberg LW, Frankenberg S, and Walker RB (in

press) Regress what on what? Paleodemographic age estimation as a calibration problem. In RR Paine (ed.): Integrating Archaeological Demography: Multidisciplinary Approaches to Prehistoric Population. Carbondale: Center for Archaeological Investigations Occasional Papers No. 24.

Lovejoy CO, Meindl RS, Pryzbeck TR, Barton TS, Heiple KG, and Kotting D (1977) Paleodemography of the Libben site, Ottawa County, Ohio. Science *198:*291–293.

Lovejoy CO, Meindl RS, Mensforth RP, and Barton TS (1985) Multifactoral determination of skeletal age at death: A method and blind tests of its accuracy. Am. J. Phys. Anthropol. *68:*1–14.

Masset CT (1989) Age estimation on the basis of cranial sutures. In MY Iscan (ed.): Age Markers in the Human Skeleton. Springfield: C.C. Thomas, pp. 71–103.

Meindl RS, and Lovejoy CO (1989) Age changes in the pelvis: Implications for paleodemography. In MY Iscan (ed.): Age Markers in the Human Skeleton. Springfield: C.C. Thomas, pp. 71–103.

Milner GR, Humpf DA, and Harpending HC (1989) Pattern matching of age-of-death distributions in paleodemographic analysis. Am. J. Phys. Anthropol. *80:* 49–58.

Newell C (1988) Methods and Models in Demography. New York: Guilford Press.

Paine RR (1989a) Model life table fitting by maximum likelihood estimation: A procedure to reconstruct paleodemographic characteristics from skeletal age distributions. Am. J. Phys. Anthropol. *79:*51–62.

Paine RR (1989b) Model life tables as a measure of bias in the Grasshopper Pueblo skeletal series. Am. Antiquity *54:*820–824.

Paine RR (1992) Population dynamics at Copán, Honduras, AD 450–1250: A study in archaeological demography. PhD Dissertation, Pennsylvania State University. Ann Arbor: University Microfilms.

Paine RR (in press) The role of uniformitarian models in osteological paleodemography. In RR Paine (ed.): Integrating Archaeological Demography: Multidisciplinary Approaches to Prehistoric Population. Carbondale: Center for Archaeological Investigations Occasional Papers No. 24.

Paine RR, and Harpending HC (in preparation) Effects of sample bias on paleodemographic reconstructions of birth.

Sattenspiel L, and Harpending HC (1983) Stable population and skeletal age. Am. Antiquity *48:*489–498.

Sullivan NC (in press) Contact period Huron demography. In RR Paine (ed.): Integrating Archaeological Demography: Multidisciplinary Approaches to Prehistoric Population. Carbondale: Center for Archaeological Investigations Occasional Papers No. 24.

Walker PL, Johnson JR, and Lambert PM (1988) Age and sex biases in the preservation of human skeletal remains. Am. J. Phys. Anthropol. *76:*183–188.

Weiss KM (1973) Demographic Models for Archeology. SAA Memoir Number 27. Washington, DC: Society for American Archaeology.

Zimmerman J, Emerson T, Willey P, Swegle M, Gregg J, Gregg P, White E, Smith C, Haberman T, Bumsted M (1981) The Crow Creek Site (39BF11) Massacre: A Preliminary Report. Omaha: Omaha District, U.S. Army Corps of Engineers.